

SURPRISING ASYMPTOTIC CONICAL STRUCTURE IN CRITICAL SAMPLE EIGEN-DIRECTIONS

BY DAN SHEN^{*,†}, HAIPENG SHEN[‡], HONGTU ZHU[§] AND J. S. MARRON[¶]

University of North Carolina at Chapel Hill

The aim of this paper is to establish several deep theoretical properties of principal component analysis for multiple-component spike covariance models. Our new results reveal a surprising asymptotic conical structure in critical sample eigendirections under the spike models with distinguishable (or indistinguishable) eigenvalues, when the sample size and/or the number of variables (or dimension) tend to infinity. The consistency of the sample eigenvectors relative to their population counterparts is determined by the ratio between the dimension and the product of the sample size with the spike size. When this ratio converges to a nonzero constant, the sample eigenvector converges to a cone, with a certain angle to its corresponding population eigenvector. In the High Dimension, Low Sample Size case, the angle between the sample eigenvector and its population counterpart converges to a limiting distribution. Several generalizations of the multi-spike covariance models are also explored, and additional theoretical results are presented.

1. Introduction. Principal Component Analysis (PCA) is one of the most important visualization and dimension reduction tools. The theoretical properties of PCA, including the sample eigenvalues, eigenvectors, and PC scores, have been widely studied in different settings, when the sample size and/or the dimension increase to infinity. For example, Anderson (1963) [1] studied such properties under the classical statistical setting with $n \rightarrow \infty$ and a fixed dimension d . Johnstone and Lu (2009) [9] explored such properties under the random matrix setting with sample size $n \rightarrow \infty$ and $d \sim n$. Jung and Marron (2009) [10] derived such properties in a High Dimension, Low Sample Size (HDLSS) context, with a fixed n and $d \rightarrow \infty$. More recently, Fan et al. (2013) [7] considered scenarios where the first few leading

^{*}Corresponding Author

[†]Partially supported by NSF grant DMS-0854908

[‡]Partially supported by NSF grants DMS-1106912 and CMMI-0800575, NIH Challenge Grant 1 RC1 DA029425-01, and the Xerox Foundation UAC Award

[§]Partially supported by NIH grants RR025747-01, P01CA142538-01, MH086633, EB005149-01 and AG033387

[¶]Partially supported by NSF grants DMS-0606577 and DMS-0854908

AMS 2000 subject classifications: Primary 62H25; Secondary 62F12

Keywords and phrases: PCA, High Dimension, Boundary Behavior, Consistency

eigenvalues increase to ∞ together with d . See additional theoretical results in [2–4, 11–15, 17, 19] and references therein.

Generally speaking, the existing results indicate that the behavior of PCA strongly depend on the relationship among three key quantities: the dimension, the sample size, and the spike sizes (the relative sizes of the population eigenvalues $\{\lambda_j\}$). For instance, Shen et al (2012) [17] systematically investigated the theoretical properties of the j -th sample eigenvector and eigenvalue as $d/(n\lambda_j) \rightarrow 0$ or ∞ . Specifically, as $d/(n\lambda_j) \rightarrow 0$, the j -th sample eigenvector converges to the corresponding population eigenvector, whereas strong inconsistency follows as $d/(n\lambda_j) \rightarrow \infty$.

An interesting open question is to investigate the asymptotic properties of PCA when $d/(n\lambda_j)$ converges to a constant $c_j \in (0, \infty)$, which is the aim of this paper. A broad theoretical framework of PCA under a broad range of cases, from the classical, through random matrix theory, and on to HDLSS, is studied here. Firstly, we show a new instance of unexpected asymptotic behavior of sample eigenvectors. Specifically, the critical sample eigenvectors lie in a right circular cone around the corresponding population eigenvectors. Although these sample eigenvectors converge to the cone, their locations within the cone are random. The angles of these cones have an increasing order, which is driven by an increasing sequence of the ratios c_j . We suggest this is as surprising as the HDLSS geometric representation results discovered by Hall et al (2005) [8], and further developed by Yata and Aoshima (2012) [19].

Secondly, we further extend the new results to the multi-spike cases where the population eigenvalues are asymptotically indistinguishable. We study the angle between the corresponding sample eigenvectors and the subspace spanned by the indistinguishable population eigenvectors. In HDLSS contexts, the cone angles are always random variables, whereas such randomness disappears when the sample size increases. We also show that in HDLSS settings, the PC scores are not consistent even when the angles between the sample eigenvectors and their population counterparts converge to 0.

Next we introduce two illustrative examples to help understand the main theoretical results in the paper, where the eigenvalues are respectively asymptotically distinguishable (Example 1.1) and indistinguishable (Example 1.2). Our theorems are applicable to a much broader class of general spike models.

EXAMPLE 1.1. (*Multiple-component spike models with distinguishable eigenvalues*) Assume that X_1, \dots, X_n are random sample vectors from a d -dimensional normal distribution $N(0, \Sigma)$, where the population eigenvalues

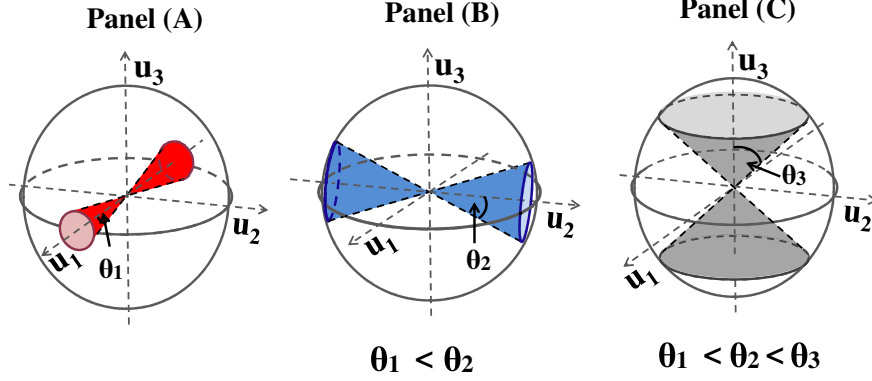


FIG 1. Geometric representation of PC directions in Example 1.1. The sphere represents the space of possible sample eigenvectors. Panel (A) shows that the first sample eigenvector tends to lie in the red cone, with the θ_1 angle. Similarly, Panels (B) and (C) show that the second and the third sample eigenvectors respectively tend to lie in the blue and the gray cones, whose angles are θ_2 and θ_3 . Note that the angle of the red cone is less than the blue cone, whose angle is again less than the gray cone.

have the following properties: as $n \rightarrow \infty$,

$$(1.1) \quad \begin{cases} \lambda_1 > \lambda_2 > \lambda_3 \gg \lambda_4 = \dots = \lambda_d = 1, \\ \frac{d}{n\lambda_j} \rightarrow c_j, \quad j = 1, 2, 3, \text{ with } 0 \leq c_1 < c_2 < c_3 \leq \infty. \end{cases}$$

In Figure 1, the sphere represents the space of all possible sample eigendirections, with the first three population eigenvectors as the coordinate axes. For this particular example, our general Theorem 3.1 suggests that

- As $n \rightarrow \infty$, the sample eigenvector \hat{u}_1 lies in the red cone, shown in Panel (A) of Fig. 1, where the angle of the cone is $\theta_1 = \arccos(\frac{1}{\sqrt{1+c_1}})$. Similarly, as $n \rightarrow \infty$, the sample eigenvectors \hat{u}_2 and \hat{u}_3 respectively lie in the blue and dark gray cones, shown in Panels (B) and (C) of Fig. 1, whereas the angles are respectively $\theta_2 = \arccos(\frac{1}{\sqrt{1+c_2}})$ and $\theta_3 = \arccos(\frac{1}{\sqrt{1+c_3}})$. Note that for $c_1 < c_2 < c_3$, we have $\theta_1 < \theta_2 < \theta_3$, as shown in Figure 1.

In addition, our Proposition 3.1 includes the two boundary cases studied by Shen et al. (2012) [17] as special cases:

- When $c_1 = c_2 = c_3 = 0$, it follows that $\theta_1 = \theta_2 = \theta_3 = 0$. This puts us in the domain of consistency [17].

- In the opposite boundary case of $c_1 = c_2 = c_3 = \infty$, we have that $\theta_1 = \theta_2 = \theta_3 = 90$ degrees. This leads to strong inconsistency [17].

Hence, our new results go well beyond the work of [17], and completely characterize the transition between consistency and strong inconsistency.

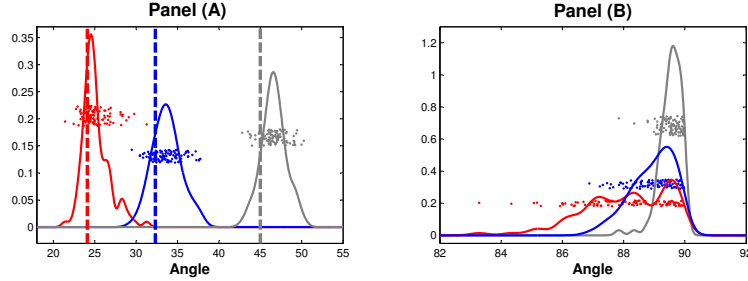


FIG 2. Example 1.1: Simulated angles between sample and population eigenvectors. Panel (A) shows realizations of angles between sample and population eigenvectors as colored dots (red is first, blue is second, gray is third). Distributions are studied using kernel density estimates, and compared with the theoretical values θ_j for $j = 1, 2, 3$, shown as dashed lines. Panel (B) studies randomness of eigen-directions within the cones shown in Figure 1, by showing the distribution of pairwise angles between realizations of the sample eigenvectors. All 3 colors are overlaid here, and all angles are very close to 90 degrees, which is very consistent with the randomness of the respective sample eigenvectors within the cones.

We investigated this theoretical convergence, using simulations, over a range of settings, with $n = 50, 100, 200, 500, 1000, 2000$, where $d/n = 50$, and $c_1 = 0.2, c_2 = 0.4, c_3 = 1$. The full sequence, illustrating this convergence, is shown in Figure A of the supplementary material [18]. Figure 2 shows the intermediate case of $n = 200$. For one data set with this distribution, we compute angles between the sample and population eigenvectors. Repeating this procedure over 100 replications, we get 100 angles for each of the first three eigenvectors, which are shown as red, blue and gray points in Panel (A). The red, blue, gray curves are the corresponding kernel density estimates. Panel (A) shows that the simulated angles are very close to the corresponding theoretical angles θ_j , $j = 1, 2, 3$, shown as dashed vertical lines.

Panel (B) in Figure 2 studies randomness of eigen-directions within the cones shown in Figure 1. We calculate pairwise angles between realizations of the sample eigenvectors for the three cones, showing angles and kernel density estimates using colors as in Panel (A) of Figure 2. All angles are very close to 90 degrees, which is consistent with randomness in high dimensions, see [8, 10, 11, 19] and the more recent work of Cai et al. (2013) [6].

In fact, the regions represented by circles in Figure 1, are actually $d - 1$ dimensional hyperspheres, so the sample eigenvectors should be thought of as $d-1$ dimensional as $d, n \rightarrow \infty$.

EXAMPLE 1.2. (Multiple-component spike models with indistinguishable eigenvalues) We again assume that X_1, \dots, X_n are random sample vectors from a d -dimensional normal distribution $N(0, \Sigma)$. Different from Example 1.1, the six leading population eigenvalues of Σ fall into three asymptotically separable pairs as follows: as $n \rightarrow \infty$

$$\begin{cases} \lambda_1 = \lambda_2 > \lambda_3 = \lambda_4 > \lambda_5 = \lambda_6 \gg \lambda_7 = \dots = \lambda_d = 1, \\ \frac{d}{n\lambda_{2j-1}} \rightarrow c_j, \quad j = 1, 2, 3, \text{ with } 0 \leq c_1 < c_2 < c_3 \leq \infty. \end{cases}$$

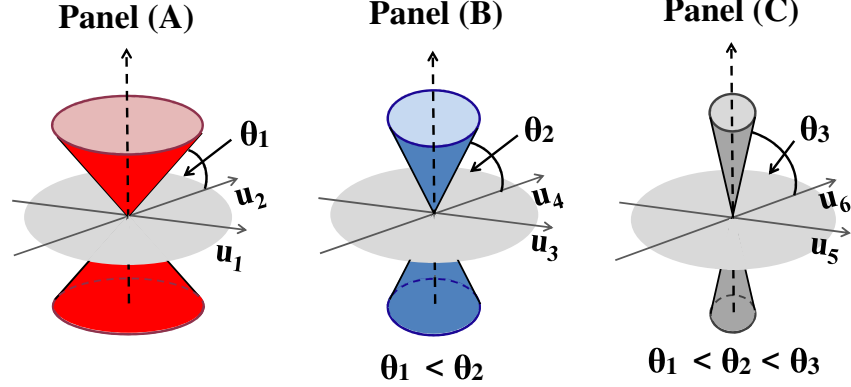


FIG 3. Example 1.2: Geometric representation of PC directions. Panel (A) shows the cone to which the first group of sample eigenvectors converge in the red. This cone has angle θ_1 with the gray subspace, generated by the first group of population eigenvectors. Similarly, Panel (B) (Panel (C)) shows the cone to which the second (third) group of sample eigenvectors converges shown as a blue (dark gray) cone, which has angle θ_2 (θ_3) with the subspace, generated by the second (third) group of population eigenvectors.

Our general Theorem 3.2, when applied to the current example, reveals the following insights:

- Panel (A) in Figure 3 shows, as a red cone, the region where the first group of sample eigenvectors \hat{u}_1 and \hat{u}_2 lie in the limit as $n \rightarrow \infty$. This has the angle $\theta_1 = \arccos(\frac{1}{\sqrt{1+c_1}})$ with the gray subspace, generated by the first group of population eigenvectors u_1 and u_2 . Similarly, Panel (B) (Panel (C)) presents, as a blue (gray) cone, the region where the

second (third) group of sample eigenvectors \hat{u}_3 and \hat{u}_4 (\hat{u}_5 and \hat{u}_6) lie in the limit as $n \rightarrow \infty$. This has the angle $\theta_2 = \arccos(\frac{1}{\sqrt{1+c_2}})$ ($\theta_3 = \arccos(\frac{1}{\sqrt{1+c_3}})$) with the subspace, generated by the second (third) group of population eigenvectors u_3 and u_4 (u_5 and u_6). Note that for $c_1 < c_2 < c_3$, we have $\theta_1 < \theta_2 < \theta_3$, as shown in Figure 3.

Furthermore, our Proposition B.1 in the supplementary document [18] considers boundary cases of our general framework, which includes the results of Shen et al. (2012) [17] as special cases:

- For $c_1 = c_2 = c_3 = 0$, it follows that $\theta_1 = \theta_2 = \theta_3 = 0$. This puts us in the domain of subspace consistency, as studied in Theorem 4.3 of [17].
- When $c_1 = c_2 = c_3 = \infty$, we have that $\theta_1 = \theta_2 = \theta_3 = 90$ degrees. This leads to strong inconsistency, as studied in Theorem 4.3 of [17].

The rest of the paper is organized as follows. Section 2 introduces the assumptions and notation relevant to the theorems in the paper. Section 3 studies the asymptotic properties of PCA for multiple spike models with distinguishable (or indistinguishable) eigenvalues as $n \rightarrow \infty$. Section 4 studies the asymptotic properties of PCA in the HDLSS contexts. Section 5 contains the technical proofs of the main theorems. Additional simulation studies and proofs can be found in the supplementary document [18].

2. Assumptions and Notation. Let X_1, \dots, X_n be random vectors from a d -dimensional normal distribution $N(\xi, \Sigma)$, where ξ is a $d \times 1$ mean vector and Σ is a $d \times d$ covariance matrix. Let $\{(\lambda_k, u_k) : k = 1, \dots, d\}$ be the eigenvalue-eigenvector pairs of Σ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$. Thus, Σ has the following eigen-decomposition

$$\Sigma = U\Lambda U^T,$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ and $U = [u_1, \dots, u_d]$. Since the relative sizes, rather than the absolute values, of the population eigenvalues affect the asymptotic properties of PCA, we assume that $\lambda_d = 1$ throughout the rest of the paper.

Let \bar{X} be the sample mean. As discussed in [16],

$$(2.1) \quad \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad \text{has the same distribution as} \quad \sum_{i=1}^{n-1} Y_i Y_i^T,$$

where Y_i are i.i.d random vectors from $N(0, \Sigma)$. It follows from (2.1) that the sample covariance matrix is location invariant. Thus, we can assume without loss of generality (WLOG):

ASSUMPTION 2.1. X_1, \dots, X_n are i.i.d random vectors from a d -dimensional normal distribution $N(0, \Sigma)$.

Denote the j th normalized population PC score vector as

$$(2.2) \quad S_j = (S_{1,j}, \dots, S_{n,j})^T = \lambda_j^{-\frac{1}{2}} (u_j^T X_1, \dots, u_j^T X_n)^T, \quad j = 1, \dots, d,$$

and define Z as the $n \times d$ random matrix as

$$(2.3) \quad Z = (z_{i,j})_{n \times d} = X^T U \Lambda^{-\frac{1}{2}},$$

where $X = [X_1, \dots, X_n]$ and $z_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, d$ are i.i.d random variables from $N(0, 1)$.

Let $\{(\hat{\lambda}_k, \hat{u}_k) : k = 1, \dots, d\}$ be the eigenvalue-eigenvector pairs of the sample covariance matrix $\hat{\Sigma} = n^{-1} X X^T$ such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_d$. Thus, $\hat{\Sigma}$ can be decomposed as

$$(2.4) \quad \hat{\Sigma} = \hat{U} \hat{\Lambda} \hat{U}^T,$$

where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_d)$ and $\hat{U} = [\hat{u}_1, \dots, \hat{u}_d]$. Note that the data matrix $n^{-\frac{1}{2}} X$ has the singular value decomposition such that $n^{-\frac{1}{2}} X = \sum_{j=1}^d \hat{\lambda}_j^{\frac{1}{2}} \hat{u}_j \hat{v}_j^T$, where $\hat{v}_j = (\hat{v}_{1,j}, \dots, \hat{v}_{n,j})^T$ for $j = 1, \dots, d$. Thus, the j th normalized sample PC score vector is given by

$$(2.5) \quad \hat{S}_j = (\hat{S}_{1,j}, \dots, \hat{S}_{n,j})^T = (\hat{v}_{1,j}, \dots, \hat{v}_{n,j})^T, \quad j = 1, \dots, d.$$

We introduce an asymptotic notation. Assume that $\{\xi_k : k = 1, \dots, \infty\}$ ($k = n$ or d) is a sequence of random variables and $\{e_k : k = 1, \dots, \infty\}$ is a sequence of constants. Denote $\xi_k = O_{\text{a.s.}}(e_k)$ if $\overline{\lim}_{k \rightarrow \infty} \left| \frac{\xi_k}{e_k} \right| \leq \zeta$ almost surely with $P(0 < \zeta < \infty) = 1$.

3. Growing sample size asymptotics. We now study asymptotic properties of PCA as $n \rightarrow \infty$. We consider multiple component spike models with distinguishable population eigenvalues in Section 3.1 and with indistinguishable eigenvalues in Section 3.2. Moreover, we vary d from the classical d fixed asymptotics, through the random matrix version with $d \sim n$, all the way to the high dimension medium sample size (HDMSS) asymptotics of Cabanski et al (2010) [5] and Yata and Aoshima (2012) [20] with $d \gg n \rightarrow \infty$.

3.1. Multiple component spike models with distinguishable eigenvalues.

We consider multiple component spike models with m dominating spikes where finite $m \in [1, n \wedge d]$. The population eigenvalues are assumed to satisfy the following two assumptions:

- $\mathcal{A1}$. As $n \rightarrow \infty$, $\lambda_1 > \dots > \lambda_m \gg \lambda_{m+1} \rightarrow \dots \rightarrow \lambda_d = 1$.
- $\mathcal{A2}$. As $n \rightarrow \infty$, $\frac{d}{n\lambda_j} \rightarrow c_j$, where $0 < c_1 < \dots < c_m < \infty$.

We first make several comments about Assumptions $\mathcal{A1}$ and $\mathcal{A2}$.

- Assumption $\mathcal{A1}$ includes two separate parts:
 - (a) The $\lambda_1 > \dots > \lambda_m$ part makes it possible to separately consider the first m principle component signals and study the corresponding asymptotic properties.
 - (b) The $\lambda_m \gg \lambda_{m+1} \rightarrow \dots \rightarrow \lambda_d = 1$ enables clear separation of the signal (contained in the first m components) from the noise (in the higher order components), which then helps to derive the asymptotic properties of the first m sample eigenvalues, eigenvectors, and PC scores.
- Assumption $\mathcal{A2}$ is the critical case, in which the *positive* information and the *negative* are of the same order. In particular, increasing n and the spike positively impacts the consistency of PCA, whereas increasing d has a negative impact.

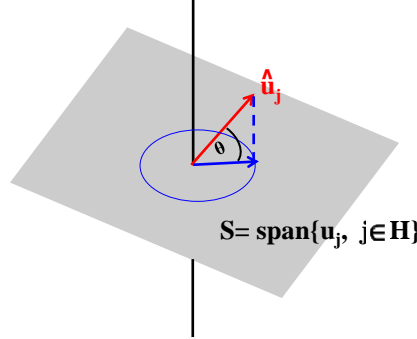


FIG 4. Angle between the sample eigenvector \hat{u}_j and the space \mathbb{S} . The blue vector is the projection of the red vector \hat{u}_j onto the space \mathbb{S} .

While the main focus of our results is the signal eigenvectors, some notation for the noise eigenvectors is also useful. According to Assumption $\mathcal{A1}$, the noise sample eigenvalues whose indices are greater than m can not

be asymptotically distinguished, so the corresponding eigenvectors should be treated as a whole. Therefore, we define the noise index set $H = \{m + 1, \dots, d\}$, and denote the space spanned by these noise eigenvectors as

$$(3.1) \quad \mathbb{S} = \text{span}\{u_j, j \in H\}.$$

For each sample eigenvector \hat{u}_j , $j \in H$, we study the angle between \hat{u}_j and the space \mathbb{S} , as defined in [10, 17] and illustrated in Figure 4, i.e. the angle between \hat{u}_j (the red vector) and its projection onto \mathbb{S} (the blue vector).

The following theorem derives the asymptotic properties of the first m sample eigenvalues and eigenvectors. In addition, the theorem also shows that, for $j = m + 1, \dots, [n \wedge d]$, the angle between \hat{u}_j and u_j goes to 90 degrees, whereas the angle between \hat{u}_j and the space \mathbb{S} goes to 0.

THEOREM 3.1. *Under Assumptions 2.1, A1, and A2, as $n \rightarrow \infty$, the sample eigenvalues satisfy*

$$(3.2) \quad \begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s.}} 1 + c_j, & 1 \leq j \leq m, \\ \frac{n\hat{\lambda}_j}{d\lambda_j} \xrightarrow{\text{a.s.}} 1, & m + 1 \leq j \leq [n \wedge d], \end{cases}$$

and the sample eigenvectors satisfy

$$(3.3) \quad \begin{cases} |\langle \hat{u}_j, u_j \rangle| \xrightarrow{\text{a.s.}} (1 + c_j)^{-\frac{1}{2}}, & 1 \leq j \leq m, \\ |\langle \hat{u}_j, u_j \rangle| = O_{\text{a.s.}} \left\{ \left(\frac{n}{d} \right)^{\frac{1}{2}} \right\}, & m + 1 \leq j \leq [n \wedge d], \\ \text{angle} \langle \hat{u}_j, \mathbb{S} \rangle \xrightarrow{\text{a.s.}} 0, & m + 1 \leq j \leq [n \wedge d]. \end{cases}$$

We now offer several remarks regarding Theorem 3.1.

REMARK 3.1. *The results of (3.2) and (3.3) suggest that, as the eigenvalue index increases, the proportional bias between the sample and population eigenvalue increases, so the angle between the sample and corresponding population eigenvectors increases. This is because larger eigenvalues (i.e. with small indices) contain more positive information, which makes the corresponding sample eigenvalues/eigenvectors less biased. These results are graphically illustrated in Figure 1 and empirically verified in Figure 2, for the specific model in Example 1.1. More empirical support is provided in the supplementary material [18].*

REMARK 3.2. *Theorem 3.1 can be extended to include the classical and random matrix cases, by allowing $c_{m_0} = 0$ for some $m_0 \leq m$, which suggests that positive information dominates in the leading m_0 spikes. Then Assumptions A1 and A2 respectively become*

A3. as $n \rightarrow \infty$, the population eigenvalues satisfy

$$\lambda_1 > \cdots > \lambda_{m_0} \gg \lambda_{m_0+1} > \cdots > \lambda_m \gg \lambda_{m+1} \rightarrow \cdots \rightarrow \lambda_d = 1.$$

A4. as $n \rightarrow \infty$, $d/(n\lambda_j) \rightarrow c_j$ for $j = 1, \dots, m$, where $0 = c_1 = \cdots = c_{m_0} < c_{m_0+1} < \cdots < c_m < \infty$.

For the classical case with fixed dimension d , $m_0 = m = d$ in Assumptions A3 and A4. For random matrix cases with $n \sim d$, $m_0 = m$ in Assumptions A3 and A4. Since $c_1 = \cdots = c_{m_0} = 0$ in Assumption A3, if the eigenvalue index is less than or equal to m_0 , the corresponding sample eigenvalues and eigenvectors are consistent. These results are summarized in the following Proposition 3.1(a).

REMARK 3.3. *Another extension of Theorem 3.1 is to allow $c_{m_0+1} = \infty$ for some $m_0 \leq m$, i.e. negative information dominates in higher-order spikes. This contains the HDMSS cases [5, 20], where $d \gg n \rightarrow \infty$. Assumption A1 then becomes Assumption A3, and Assumption A2 becomes*

A5. as $n \rightarrow \infty$, $d/(n\lambda_j) \rightarrow c_j$ for $j = 1, \dots, m$, where $0 < c_1 < \cdots < c_{m_0} < c_{m_0+1} = \cdots = c_m = \infty$.

Since $c_{m_0+1} = \cdots = c_m = \infty$, for index $j \geq m_0 + 1$, the proportional error between the sample and population eigenvalues goes to infinity, and the angle between the corresponding sample and population eigenvectors converges to 90 degrees. These results are summarized in Proposition 3.1(b).

PROPOSITION 3.1. (a) *Under Assumptions 2.1, A3 and A4, the sample eigenvalues and eigenvectors satisfy*

$$\hat{\lambda}_j/\lambda_j \quad \text{and} \quad |\langle \hat{u}_j, u_j \rangle| \xrightarrow{\text{a.s.}} 1, \quad 1 \leq j \leq m_0,$$

and the properties of the other sample eigenvalues and eigenvectors remain the same as in Theorem 3.1.

(b) *Let $H = \{m_0 + 1, \dots, d\}$ and define \mathbb{S} as in (3.1). If Assumption A4 in (a) is replaced by Assumption A5, the sample eigenvalues satisfy*

$$n\hat{\lambda}_j/d \xrightarrow{\text{a.s.}} 1, \quad m_0 + 1 \leq j \leq m,$$

and the sample eigenvectors satisfy

$$\begin{cases} |\langle \hat{u}_j, u_j \rangle| = O_{\text{a.s.}} \left\{ \left(\frac{n\lambda_j}{d} \right)^{\frac{1}{2}} \right\}, & m_0 \leq j \leq [n \wedge d]; \\ \text{angle} < \hat{u}_j, \mathbb{S} \xrightarrow{\text{a.s.}} 0, \end{cases}$$

the properties of the other sample eigenvalues and eigenvectors remain the same as in Theorem 3.1.

(c) In addition, if Assumption $\mathcal{A4}$ in (a) is strengthened to $d/\lambda_{m_0} \rightarrow 0$, then the sample PC scores satisfy

$$\left| \frac{\hat{S}_{i,j}}{S_{i,j}} \right| \xrightarrow{a.s} 1, \quad i = 1, \dots, n, \quad j = 1, \dots, m_0.$$

3.2. Multiple component spike models with indistinguishable eigenvalues.

We now consider spike models with the m leading eigenvalues being grouped into $r(\geq 1)$ tiers, each of which contains eigenvalues that are either the same or have the same limit. The eigenvalues within different tiers have different limits. Specifically, the first m eigenvalues are grouped into r tiers, in which there are q_k eigenvalues in the k th tier such that $\sum_{l=1}^r q_l = m$. Define $q_0 = 0$, $q_{r+1} = d - \sum_{l=1}^r q_l$, and the index set of the eigenvalues in the k th tier as

$$(3.4) \quad H_k = \left\{ \sum_{l=0}^{k-1} q_l + 1, \sum_{l=0}^{k-1} q_l + 2, \dots, \sum_{l=0}^{k-1} q_l + q_k \right\}, \quad k = 1, \dots, r+1.$$

We make the following assumptions on the tiered eigenvalues:

$\mathcal{B1}$. The eigenvalues in the k th tier have the same limit $\delta_k(> 0)$:

$$\lim_{n \rightarrow \infty} \frac{\lambda_j}{\delta_k} = 1, \quad j \in H_k, \quad k = 1, \dots, r.$$

$\mathcal{B2}$. The eigenvalues in different tiers have different limits:

$$\text{as } n \rightarrow \infty, \quad \delta_1 > \dots > \delta_r \gg \lambda_{m+1} \rightarrow \dots \rightarrow \lambda_d = 1.$$

$\mathcal{B3}$. The ratio between the dimension and the product of the sample size with eigenvalues in the same tier converges to a constant:

$$\text{as } n \rightarrow \infty, \quad \frac{d}{n\delta_k} \rightarrow c_k, \quad \text{with } 0 < c_1 < \dots < c_r < \infty.$$

Assumptions $\mathcal{B2}$ and $\mathcal{B3}$ are natural extensions of Assumptions $\mathcal{A1}$ and $\mathcal{A2}$. In Assumption $\mathcal{B2}$, the signal contained in the first r tiers of eigenvalues is well separated from the noise, and hence the asymptotic properties of the sample eigenvalues and eigenvectors in the first r tiers can be obtained. Assumption $\mathcal{B3}$ suggests that the positive information (sample size and spike size) and the negative information (dimension) are of the same order.

Since the sample eigenvalues within the same tier can not be asymptotically identified, the corresponding sample eigenvectors are indistinguishable.

For $j \in H_k$, in order to study the asymptotic properties of the sample eigenvector \hat{u}_j , we consider the angle between \hat{u}_j and the subspace spanned by the population eigenvectors u_j in the same tier, defined as

$$(3.5) \quad \mathbb{S}_k = \text{span}\{u_j, j \in H_k\}.$$

Our theoretical results are summarized in the following theorem.

THEOREM 3.2. *Under Assumptions 2.1, B1, B2 and B3, as $n \rightarrow \infty$, the sample eigenvalues satisfy*

$$(3.6) \quad \begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s.}} 1 + c_k, & j \in H_k, k = 1, \dots, r, \\ \frac{n\hat{\lambda}_j}{d\lambda_j} \xrightarrow{\text{a.s.}} 1, & m+1 \leq j \leq [n \wedge d], \end{cases}$$

and the sample eigenvectors satisfy

$$(3.7) \quad \begin{cases} \text{angle} < \hat{u}_j, \mathbb{S}_k > \xrightarrow{\text{a.s.}} \arccos \left\{ (1 + c_k)^{-\frac{1}{2}} \right\}, & j \in H_k, k = 1, \dots, r, \\ |\langle \hat{u}_j, u_j \rangle| = O_{\text{a.s.}} \left\{ \left(\frac{n}{d} \right)^{\frac{1}{2}} \right\}, & m+1 \leq j \leq [n \wedge d], \\ \text{angle} < \hat{u}_j, \mathbb{S}_{r+1} > \xrightarrow{\text{a.s.}} 1, & m+1 \leq j \leq [n \wedge d]. \end{cases}$$

Theorem 3.2 is an extension of Theorem 3.1. For higher-order eigenvalues, the sample eigenvalues are more biased, while the angles between the sample eigenvectors and the subspaces spanned by their population counterparts in the same tiers are larger. See Figure 3 for an illustration of the specific model considered in Example 1.2. Theorem 3.2 can be extended to cover the classical, random matrix, and HDMSS cases, which is done in Section B of the supplementary material [18].

4. High dimension, low sample size asymptotics. We now study the asymptotic properties of PCA in the HDLSS context. In this case, the ratios between the sample eigenvalues and their population counterparts converge to non degenerate random variables, as do the angles between the sample eigenvectors and the space spanned by the corresponding population eigenvectors. This phenomenon of random limits does not exist when n increases to ∞ as shown in Section 3.

Since the sample size is fixed, we can not distinguish the two types of spike models considered respectively in Sections 3.1 and 3.2. Hence, we merge the model assumptions there into the following corresponding assumptions:

C1. For fixed n , as $d \rightarrow \infty$, $\lambda_1 \geq \dots \geq \lambda_m \gg \lambda_{m+1} \rightarrow \dots \rightarrow \lambda_d = 1$.

$\mathcal{C}2$. For fixed n , as $d \rightarrow \infty$,

$$\frac{d}{n\lambda_j} \rightarrow c_j, \quad \text{with } 0 < c_1 \leq \cdots \leq c_m < \infty.$$

In particular, Assumption $\mathcal{C}1$ is parallel to Assumptions $\mathcal{A}1$, $\mathcal{B}1$ and $\mathcal{B}2$, while Assumption $\mathcal{C}2$ corresponds to Assumptions $\mathcal{A}2$ and $\mathcal{B}3$.

As stated below in Theorem 4.1, the sample eigenvalues and eigenvectors converge to non-degenerate random variables rather than constants. We define several quantities in order to describe the limiting random variables. Define the $m \times d$ matrix

$$\mathbb{M} = [\mathbb{C}, 0_{m \times (d-m)}]_{m \times d},$$

where $\mathbb{C} = \text{diag}\{c_1^{-1/2}, \dots, c_m^{-1/2}\}$ is an $m \times m$ diagonal matrix and $0_{m \times (d-m)}$ is the $m \times (d-m)$ zero matrix. In addition, define the random matrix \mathcal{W} as

$$(4.1) \quad \mathcal{W} = \mathbb{M} Z^T Z \mathbb{M}^T,$$

where Z is defined in (2.3). The eigenvalues of the random matrix \mathcal{W} appear in the random limits of Theorem 4.1, as in (4.2) and (4.3).

Given the fixed sample size, the sample eigenvalues can not be asymptotically distinguished, nor can the corresponding sample eigenvectors. To study the asymptotic behavior of the sample eigenvectors, we need to consider the space \mathbb{S}_k spanned by the corresponding population eigenvectors, as defined in (3.5), with the two index sets being $H_1 = \{1, \dots, m\}$ and $H_2 = \{m+1, \dots, d\}$.

We are now ready to state the main theorem in the HDLSS contexts.

THEOREM 4.1. *Under Assumptions 2.1, $\mathcal{C}1$ and $\mathcal{C}2$, for fixed n , as $d \rightarrow \infty$, the sample eigenvalues satisfy*

$$(4.2) \quad \begin{cases} \frac{\hat{\lambda}_j}{\lambda_j} \xrightarrow{\text{a.s.}} \frac{c_j}{n} \lambda_j(\mathcal{W}) + c_j, & 1 \leq j \leq m, \\ \frac{n\lambda_j}{d\lambda_j} \xrightarrow{\text{a.s.}} 1, & m+1 \leq j \leq n, \end{cases}$$

where \mathcal{W} is defined in (4.1), and the sample eigenvectors satisfy

$$(4.3) \quad \begin{cases} \text{angle} < \hat{u}_j, \mathbb{S}_1 > \xrightarrow{\text{a.s.}} \arccos \left\{ \left(1 + \frac{n}{\lambda_j(\mathcal{W})} \right)^{-\frac{1}{2}} \right\}, & 1 \leq j \leq m, \\ |\langle \hat{u}_j, u_j \rangle| = O_{\text{a.s.}}(d^{-\frac{1}{2}}), & m+1 \leq j \leq n, \\ \text{angle} < \hat{u}_j, \mathbb{S}_2 > \xrightarrow{\text{a.s.}} 1, & m+1 \leq j \leq n. \end{cases}$$

Three remarks are offered below regarding Theorem 4.1.

REMARK 4.1. *If $m = 1$ in Theorem 4.1, i.e. for single-component spike models, then the first sample eigenvalue and eigenvector satisfy*

$$\begin{cases} \frac{\hat{\lambda}_1}{\lambda_1} \xrightarrow{a.s.} \frac{\chi_n^2}{n} + c_1, \\ |\langle \hat{u}_1, u_1 \rangle| \xrightarrow{a.s.} \left(1 + \frac{nc_1}{\chi_n^2}\right)^{-\frac{1}{2}}, \end{cases}$$

where χ_n^2 is the Chi-square distribution with n degrees of freedom. This result is consistent with Theorem 1 of Jung et al. (2012) [11].

REMARK 4.2. *For $1 \leq j \leq m$, as the relative size of the eigenvalue decreases, the angle between \hat{u}_j and \mathbb{S}_1 increases. However, this phenomenon is not as strong as in the growing sample size settings studied in Section 3, where the sample eigenvectors can be separately studied, and the corresponding angles have a non-random increasing order.*

REMARK 4.3. *Assumption C2 can be relaxed to include boundary cases, in which there exists an integer $m_0 \in [1, m]$ such that $c_{m_0} = 0$, i.e. positive information dominates in the leading m_0 spikes; or $c_{m_0+1} = \infty$, i.e. negative information dominates in the remaining high-order spikes. These theoretical results are presented in Section C of the supplementary material [18].*

5. Proofs. We now provide some proofs of our theorems as $n \rightarrow \infty$. For the sake of space, we only present detailed proof for the properties of the sample eigenvectors here, which is the most challenging part. In contrast to showing consistency or inconsistency of the sample eigenvector, this proof requires precise calculation of the degree of inconsistency, i.e. the limiting angles between the sample and population eigenvectors. We relegate the derivations regarding the sample eigenvalues to Section D of the supplementary material [18], which also contains proofs of Proposition 3.1, Theorem 4.1, as well as extensions of Theorems 3.2 and 4.1.

The critical ideas of the proof are to first partition the sample eigenvector matrix \hat{U} into sub-matrices, corresponding to the group index H_k . Then through careful analysis, we explore the connections between sample eigenvectors and eigenvalues and then use the sample eigenvalue properties to study the asymptotic properties of the sample eigenvectors.

WLOG, we assume that $\lambda_{m+1} = \cdots = \lambda_d = 1$. Due to the invariance property of the angle between the sample and population eigenvectors, see Shen et al. (2012) [17], we assume WLOG that the population eigenvectors $u_j = e_j$, $j = 1, \dots, d$, where the j -th component of e_j equals 1 and the

rest are zero. It follows that the inner product between the sample and population eigenvectors satisfies

$$|\langle \hat{u}_j, u_j \rangle|^2 = |\langle \hat{u}_j, e_j \rangle|^2 = \hat{u}_{j,j}^2,$$

and the angle between the sample eigenvector and the corresponding population subspace \mathbb{S}_k in (3.5) satisfies

$$(5.1) \quad (\cos [\text{angle}(\hat{u}_j, \mathbb{S}_k)])^2 = \sum_{l \in H_k} \hat{u}_{l,j}^2, \quad k = 1, \dots, r+1.$$

The population eigenvalues are grouped into $r+1$ tiers and H_k in (3.4) is the index set of the eigenvalues in the k th tier. Define

$$\hat{U}_{k,l} = (\hat{u}_{i,j})_{i \in H_k, j \in H_l}, \quad 1 \leq k, l \leq r+1.$$

Then, the sample eigenvector matrix \hat{U} can be expressed as:

$$(5.2) \quad \hat{U} = [\hat{u}_1, \hat{u}_2, \dots, \hat{u}_d] = \begin{pmatrix} \hat{U}_{1,1} & \hat{U}_{1,2} & \cdots & \hat{U}_{1,r+1} \\ \hat{U}_{2,1} & \hat{U}_{2,2} & \cdots & \hat{U}_{2,r+1} \\ \vdots & \vdots & & \vdots \\ \hat{U}_{r+1,1} & \hat{U}_{r+1,2} & \cdots & \hat{U}_{r+1,r+1} \end{pmatrix}.$$

The proof of the asymptotic properties of the sample eigenvectors (3.7) depends on the asymptotic properties of the sample eigenvalues, as stated in (3.6) of Theorem 3.2, which are derived in the supplementary material [18]. The following proof considers two groups of sample eigenvectors separately. Section 5.1 obtains the asymptotic properties for the sample eigenvectors whose index is greater than m . Section 5.2 derives asymptotic properties for the sample eigenvectors whose index is less than or equal to m .

5.1. *Asymptotic properties of the sample eigenvectors \hat{u}_j with $j > m$.* We derive the asymptotic properties through the following two steps:

- First, we show that as $n \rightarrow \infty$, the angle between \hat{u}_j and u_j converges to 90 degrees:

$$(5.3) \quad |\langle \hat{u}_j, u_j \rangle|^2 = \hat{u}_{j,j}^2 = O_{a.s} \left(\frac{n}{d} \right), \quad j = m+1, \dots, [n \wedge d].$$

- Then, we show that as $n \rightarrow \infty$, the angle between \hat{u}_j and the corresponding subspace \mathbb{S}_{r+1} converges to 0, where \mathbb{S}_{r+1} is defined as in (3.5):

$$(5.4) \quad \text{angle} \langle \hat{u}_j, \mathbb{S}_{r+1} \rangle \xrightarrow{a.s} 0, \quad j = m+1, \dots, [n \wedge d].$$

We now provide the proof for the first step. Denote $W = \Lambda^{-\frac{1}{2}} \hat{U} \hat{\Lambda}^{\frac{1}{2}}$, where \hat{U} is the sample eigenvector matrix and $\hat{\Lambda}$ is the sample eigenvalue matrix defined in (2.4). It follows from (2.3) and (2.4) that $WW^T = \frac{1}{n}ZZ^T$, where Z is defined in (2.3). Considering the k -th diagonal entry of the two equivalent matrices WW^T and $\frac{1}{n}ZZ^T$, and noting that $w_{k,j} = \lambda_k^{-\frac{1}{2}} \hat{\lambda}_j^{\frac{1}{2}} \hat{u}_{k,j}$, it follows that

$$(5.5) \quad \lambda_k^{-1} \sum_{j=1}^d \hat{\lambda}_j \hat{u}_{k,j}^2 = \sum_{j=1}^d w_{k,j}^2 = \frac{1}{n} \sum_{i=1}^n z_{i,k}^2, \quad k = 1, \dots, d.$$

In addition, note that $\frac{1}{n} \sum_{i=1}^n z_{i,k}^2 \xrightarrow{\text{a.s.}} 1$, as $n \rightarrow \infty$, and $\hat{\lambda}_j = 0$ for $j > [n \wedge d]$. Combining the above with (5.5), we obtain that

$$(5.6) \quad \sum_{l=1}^r \sum_{j \in H_l} \lambda_k^{-1} \hat{\lambda}_j \hat{u}_{k,j}^2 + \sum_{j=m+1}^{[n \wedge d]} \lambda_k^{-1} \hat{\lambda}_j \hat{u}_{k,j}^2 \xrightarrow{\text{a.s.}} 1, \quad k = 1, \dots, d.$$

Furthermore, it follows from (5.6) that as $n \rightarrow \infty$,

$$(5.7) \quad \hat{u}_{j,j}^2 \leq \frac{\lambda_j}{\hat{\lambda}_j}, \quad j = m+1, \dots, [n \wedge d],$$

which, together with the asymptotic properties of the sample eigenvalues (3.6), yields (5.3).

We then move on to prove the second step. According to (5.1), we need to show that

$$(5.8) \quad \sum_{k=m+1}^d \hat{u}_{k,j}^2 \xrightarrow{\text{a.s.}} 1, \quad j = m+1, \dots, [n \wedge d].$$

The non-zero k -th diagonal entry of $W^T W$ is between its smallest and largest eigenvalues. Since $W^T W$ shares the same non-zero eigenvalues as $\frac{1}{n} Z^T Z$, it follows that for $j = 1, \dots, [n \wedge d]$,

$$(5.9) \quad \lambda_{\min}(\frac{1}{n} Z^T Z) \leq \hat{\lambda}_j \sum_{k=1}^d \lambda_k^{-1} \hat{u}_{k,j}^2 = \sum_{k=1}^d w_{k,j}^2 \leq \lambda_{\max}(\frac{1}{n} Z^T Z),$$

which yields that, for $j = m+1, \dots, [n \wedge d]$,

$$(5.10) \quad \frac{\lambda_j}{\hat{\lambda}_j} \lambda_{\min}(\frac{1}{n} Z^T Z) \leq \sum_{k=1}^d \lambda_j \lambda_k^{-1} \hat{u}_{k,j}^2 \leq \frac{\lambda_j}{\hat{\lambda}_j} \lambda_{\max}(\frac{1}{n} Z^T Z).$$

According to Lemma D.1 in the supplementary material [18] and the asymptotic properties of the sample eigenvalues (3.6), we have that, for $j = m + 1, \dots, [n \wedge d]$,

$$(5.11) \quad \frac{\lambda_j}{\hat{\lambda}_j} \lambda_{\min} \left(\frac{1}{n} Z^T Z \right) \quad \text{and} \quad \frac{\lambda_j}{\hat{\lambda}_j} \lambda_{\max} \left(\frac{1}{n} Z^T Z \right) \xrightarrow{a.s.} 1.$$

In addition, it follows from Assumption $\mathcal{B}2$ that, for $j = m + 1, \dots, [n \wedge d]$,

$$(5.12) \quad \begin{cases} \lambda_j \lambda_k^{-1} \rightarrow 0, & k = 1, \dots, m, \\ \lambda_j \lambda_k^{-1} \rightarrow 1, & k = m + 1, \dots, d. \end{cases}$$

Combining (5.10), (5.11), and (5.12), we have (5.8), which further leads to (5.4).

5.2. Asymptotic properties of the sample eigenvectors \hat{u}_j with $j \in [1, m]$. We need to prove that, for $j = 1, \dots, m$, the angle between the sample eigenvector \hat{u}_j and the corresponding population subspace \mathbb{S}_l , $j \in H_l$, converges to $\arccos(\frac{1}{\sqrt{1+c_l}})$, $l = 1, \dots, r$. According to (5.1), we only need to show that

$$(5.13) \quad \sum_{k \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{1}{1+c_l}, \quad j \in H_l, \quad l = 1, \dots, r.$$

Below, we provide the detailed proof of (5.13) for $l = 1$, and briefly illustrate how repeating the same procedure can lead to (5.13) for $l > 2$.

In order to show (5.13) for $l = 1$, we need the following lemma about the asymptotic properties of the eigenvector matrix \hat{U} in (5.2):

LEMMA 5.1. *Under Assumptions in Theorem 3.2 and as $n \rightarrow \infty$, the rows of the eigenvector matrix \hat{U} satisfy*

$$(5.14) \quad \sum_{l=1}^r (1+c_l) c_h c_l^{-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 1, \quad k \in H_h, \quad h = 1, \dots, r,$$

and the columns of the eigenvector matrix \hat{U} satisfy

$$(5.15) \quad \sum_{h=1}^r \sum_{k \in H_h} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{1}{1+c_l}, \quad j \in H_l, \quad l = 1, \dots, r.$$

In addition, we also have

$$(5.16) \quad \sum_{l=1}^r (1+c_l) \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 1, \quad k \in H_1.$$

Lemma 5.1 is proven in Section D.3.3 of the supplementary material [18]. We now show how to use Lemma 5.1 to prove (5.13) for $l = 1$. Let $h = 1$ in (5.14), and then we have that

$$(5.17) \quad \sum_{l=1}^r (1 + c_l) c_l c_l^{-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 1, \quad k \in H_1.$$

Note that $c_1 c_l^{-1} < 1$ for $l > 1$, and comparing (5.16) with (5.17), we get that

$$(5.18) \quad \sum_{l=2}^r \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 0, \quad \sum_{j \in H_1} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{1}{1 + c_1}, \quad k \in H_1,$$

which then yields that

$$(5.19) \quad \sum_{k \in H_1} \sum_{j \in H_1} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{q_1}{1 + c_1},$$

where q_1 is the number of eigenvalues in H_1 (3.4). Summing over $j \in H_1$ in (5.15), we have that

$$(5.20) \quad \sum_{h=1}^r \sum_{k \in H_h} \sum_{j \in H_1} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{q_1}{1 + c_1}.$$

It follows from (5.19) and (5.20) that

$$(5.21) \quad \sum_{h=2}^r \sum_{k \in H_h} \sum_{j \in H_1} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 0,$$

which, together with (5.15) for $l = 1$, yields

$$\sum_{k \in H_1} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{1}{1 + c_1}, \quad j \in H_1.$$

which is (5.13) for $l = 1$.

We now prove (5.13) for $l = 2, \dots, r$. Note that

- it follows from (5.21) that (5.14) becomes

$$(5.22) \quad \sum_{l=2}^r (1 + c_l) c_h c_l^{-1} \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 1, \quad k \in H_h, \quad h = 2, \dots, r.$$

- it follows from (5.18) that (5.15) becomes

$$(5.23) \quad \sum_{h=2}^r \sum_{k \in H_h} \hat{u}_{k,j}^2 \xrightarrow{a.s.} \frac{1}{1 + c_l}, \quad j \in H_l, \quad l = 2, \dots, r.$$

- similar to (5.16), we have

$$(5.24) \quad \sum_{l=2}^r (1 + c_l) \sum_{j \in H_l} \hat{u}_{k,j}^2 \xrightarrow{a.s.} 1, \quad k \in H_2.$$

Finally, combining (5.22), (5.23) and (5.24), we can prove (5.13) for $l = 2$. We can repeat the same procedure for $l = 3, \dots, r$.

SUPPLEMENTARY MATERIAL

Simulations and proofs

(<http://www.unc.edu/~dshen/BBPCA/BBPCASupplement.pdf>). The supplementary material contains additional simulation results that empirically verify the theoretical convergence of the angles between sample eigenvectors and their population counterparts, reported in our theorems. We also provide detailed proofs for our theorems and their extensions under both the growing sample size and HDLSS contexts.

References.

- [1] ANDERSON, T. (1963). Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics* **34** 122–148.
- [2] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability* **33** 1643–1697.
- [3] BAIK, J. and SILVERSTEIN, J. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis* **97** 1382–1408.
- [4] BENAYCH-GEORGES, F. and NADAKUDITI, R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* **227** 494–521.
- [5] CABANSKI, C., QI, Y., YIN, X., BAIR, E., HAYWARD, M., FAN, C., LI, J., WILKERSON, M., MARRON, J., PEROU, C. and HAYES, D. (2010). SWISS MADE: standardized within class sum of squares to evaluate methodologies and dataset elements. *PloS One* **5** e9905.
- [6] CAI, T., FAN, J. and JIANG, T. (2013). Distributions of angles in random packing on spheres. *Technical Report*.
- [7] FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B* **75** 1–44.
- [8] HALL, P., MARRON, J. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B* **67** 427–444.

- [9] JOHNSTONE, I. and LU, A. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* **104** 682–693.
- [10] JUNG, S. and MARRON, J. (2009). PCA consistency in high dimension, low sample size context. *The Annals of Statistics* **37** 4104–4130.
- [11] JUNG, S., SEN, A. and MARRON, J. (2012). Boundary behavior in high dimension, low sample size asymptotics of PCA. *Journal of Multivariate Analysis* **109** 190–203.
- [12] LEE, S., ZOU, F. and WRIGHT, F. A. (2010). Convergence and prediction of principal component scores in high-dimensional settings. *The Annals of Statistics* **38** 3605–3629.
- [13] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics* **36** 2791–2817.
- [14] ONATSKI, A. (2006). Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak. *Manuscript, Columbia University*.
- [15] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* **17** 1617–1642.
- [16] PAUL, D. and JOHNSTONE, I. (2007). Augmented sparse principal component analysis for high-dimensional data. *Technical Report, UC Davis*.
- [17] SHEN, D., SHEN, H. and MARRON, J. (2012). A general framework for consistency of principal component analysis. *arXiv preprint arXiv:1211.2671*.
- [18] SHEN, D., SHEN, H., ZHU, H. and MARRON, J. (2013). Surprising asymptotic conical structure in critical sample eigen-directions: supplementary materials. *Available online at <http://www.unc.edu/~dshen/BBPCA/BBPCASupplement.pdf>*.
- [19] YATA, K. and AOSHIMA, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis* **105** 193–215.
- [20] YATA, K. and AOSHIMA, M. (2012). Inference on high-dimensional mean vectors with fewer observations than the dimension. *Methodology and Computing in Applied Probability* 1–18.

DEPARTMENT OF BIOSTATISTICS
 UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
 CHAPEL HILL, NC 27599
 E-MAIL: dshen@email.unc.edu
htzhu@email.unc.edu

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
 UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
 CHAPEL HILL, NC 27599
 E-MAIL: haipeng@email.unc.edu
marron@email.unc.edu